PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04N 7/26

A1

(11) International Publication Number: WO 99/33274

(43) International Publication Date: 1 July 1999 (01.07.99)

(21) International Application Number: PCT/US98/26984

(22) International Filing Date: 18 December 1998 (18.12.98)

(30) Priority Data:
60/068,331 19 December 1997 (19.12.97) US

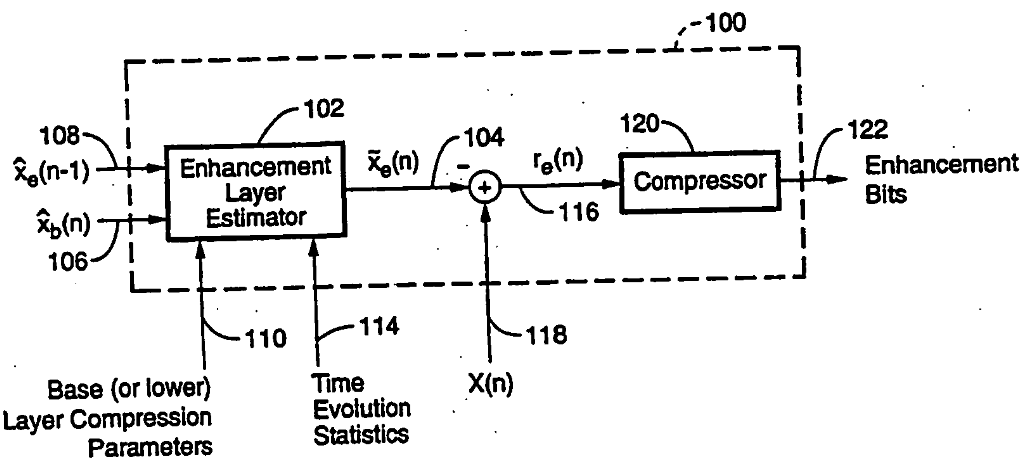
(71)(72) Applicant and Inventor: ROSE, Kenneth [US/US]; 130 Alpine Drive, Goleta, CA 93117 (US).

(74) Agent: O'BANION, John, P.; Gerber, Ritchey & O'Banion LLP, Suite 1550, 400 Capitol Mall, Sacramento, CA 95814 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SCALABLE PREDICTIVE CODING METHOD AND APPARATUS



(57) Abstract

A scalable predictive coder in which the current frame of data is predicted at the enhancement–layer by processing and combining the reconstructed signal at: (i) the current base–layer (or lower layers) frame; and (ii) the previous enhancement–layer frame. The combining rule takes into account the compressed prediction error of the base–layer, and the parameters used for its compression.

SCALABLE PREDICTIVE CODING METHOD AND APPARATUS

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention pertains generally to data compression methods and systems, and more particularly to an efficient scalable predictive coding method and system where most or all of the information available to the enhancement-layer is exploited to improve the quality of the prediction.

2. Description of the Background Art

The following publications which are referenced herein using numbers in square brackets (e.g., [1]) are incorporated herein by reference:

[1] D. Wilson and M. Ghanbari, "Transmission of SNR scalable two layer MPEG-2 coded video through ATM networks," *Proc. 7th International Workshop on Packet Video*, pp. 185-189, Brisbane Australia, March 1996.

[2] B. Girod, U. Horn, and B. Belzer, "Scalable video coding with multiscale motion compensation and unequal error protection," In Y. Wang, S. Panwar, S.-P. Kim, and H. L. Bertoni, editors, *Multimedia Communications and Video Coding*, pp. 475-482, New York: Plenum Press, 1996.

[3] B. G. Haskell, A. Puri, and A. N. Netravali, *Digital video: an introduction to MPEG-2*. New York: Chapman and Hall, International Thomson Pub., 1997.

[4] *Draft text of H.263*, Version 2 (H.263+).

[5] T. K. Tan, K. K. Pang, and K. N. Ngan, "A frequency scalable coding scheme employing pyramid and subband techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 203-207, April 1994.

[6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Press, 1992.

Many applications require data, such as video, to be simultaneously decodable at a variety of rates. Examples include applications involving broadcast over differing channels, multicast in a complex network where the channels/links dictate the feasible bit rate for each user, the co-existence of receivers of different complexity (and cost), and time-varying channels. An associated compression technique is "scalable" if it offers a variety of decoding rates using the same basic

access to both bit-streams and produces an enhanced reconstruction 28, $\hat{x}_e(n)$. The reconstruction frames that are available at the decoder are used to predict or estimate the current frame. Note that ED 26 has access to both bit streams and hence it effectively has access to both the reconstruction frame at the base layer, $\hat{x}_b(n)$, and the previous reconstructed frame at the enhancement layer $\hat{x}_e(n-1)$, while BD 22 has only access to the previous reconstructed frame at the base layer, $\hat{x}_b(n-1)$, which is stored in the memory within BD. In the case of a scalable coding system with multiple enhancement layers, an enhancement layer decoder may have access to the reconstruction frames from lower enhancement layers as well as from the base layer. The prediction loop (internal to the operation of BD as in any predictive coding system but not shown in the figure) in this configuration causes severe difficulties in the design of scalable coding. Accordingly, a number of approaches to scalable coding have been developed. These include,

(1) The standard approach: At the base layer, BE 14 compresses the residual $r_b(n) = x(n) - P[\hat{x}_b(n-1)]$, where P denotes the predictor (e.g., motion compensator in the case of video coding). Note that for notational simplicity we assume first-order prediction, but in general several previous frames may be used. BD 22 produces the reconstruction $\hat{x}_b(n) = P[\hat{x}_b(n-1)] + \hat{r}_b(n)$, where $\hat{r}_b(n)$ is the compressed-reconstructed residual. At the enhancement-layer, EE 18 compresses the base layer's reconstruction error $r_e^{(1)} = x(n) - \hat{x}_b(n) = x(n) - P[\hat{x}_b(n-1)] - \hat{r}_b(n)$. The enhancement-layer reconstruction is $\hat{x}_e(n) = \hat{x}_b(n) + \hat{r}_e^{(1)}(n) = P[\hat{x}_b(n-1)] + \hat{r}_b(n) + \hat{r}_e^{(1)}(n)$. See, e.g., [1]. A deficiency of this approach is that no advantage is taken of the potentially superior prediction due to the availability of $\hat{x}_e(n-1)$ at the ED 26.

(2) The separate coding approach: BE 14 compresses $r_b(n)$ as above, but EE 18 compresses the "enhancement-only" prediction error $r_e^{(2)} = x(n) - P[\hat{x}_e(n-1)]$ directly. The enhancement-layer reconstruction is $\hat{x}_e(n) = P[\hat{x}_e(n-1)] + \hat{r}_e^{(2)}(n)$. A deficiency of this approach is that, while the approach takes advantage of information available only to the enhancement-layer, it does not exploit the knowledge of $\hat{r}_b(n)$ which is also available at the enhancement-layer. The two layers are, in fact, separately encoded except for

and their impact as described in the Background of the Invention. This difficulty may explain why existing known methods exclusively use one of these information sources at any given time. These methods will be generally referred to here as switching techniques (which include as a special case the exclusive use of one of the information sources at all times). Additionally, the invention optionally includes a special enhancement-layer synchronization mode for the case where the communication rate for a given receiver is time varying (e.g., in mobile communications). This mode may be applied periodically to allow the receiver to upgrade to enhancement–layer performance even though it does not have prior enhancement-layer reconstructed frames.

An object of the invention is to achieve efficient scalability of predictive coding.

Another object of the invention is to provide a method and system for scalable predictive coding that is applicable to typical or standard video and audio compression.

Another object of the invention is to provide a scalable predictive coding method and system in which all or most of the information available at an enhancement-layer is exploited to improve the quality of the prediction.

Further objects and advantages of the invention will be brought out in the following portions of the specification, wherein the detailed description is for the purpose of fully disclosing preferred embodiments of the invention without placing limitations thereon.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more fully understood by reference to the following drawings which are for illustrative purposes only:

FIG. 1 is functional block diagram of a conventional two-layer scalable predicting coding system.

FIG. 2 is a functional block diagram of an enhancement layer encoder of a scalable predictive coding system in accordance with the present invention.

FIG. 3 is a functional block diagram of a base layer reconstruction module according to the present invention.

being estimated or predicted by the method in each particular application.

Referring first to FIG. 2, a functional block diagram of an enhancement layer encoder of a scalable predictive coding system in accordance with the present invention is shown. In the enhancement layer encoder 100 of the present invention, an enhancement layer estimator (ELE) 102 computes a new predicted frame 104, $\tilde{x}_e(n)$, by combining information from the reconstruction frame 106 at the base layer, $\hat{x}_b(n)$ and from the previous reconstructed frame 108 at the enhancement layer $\hat{x}_e(n-1)$. Note that first order prediction is described for notational simplicity but several previous frames may be used. The combining rule depends on any or all of, but not limited to, the following parameters: the compression parameters 110 of the base layer (such as quantization step and threshold, and the quantized base-layer residual 112, $\hat{r}_b(n)$, (see FIG. 3)), and the statistical parameters 114 of the time evolution of the frames (such as inter-frame correlation coefficients and variance). The statistical parameters may be either estimated off-line from training data, or estimated on-line by an adaptive estimator which tracks variation in the signal statistics based on either the original signal (in which case the parameters need to be transmitted to the decoder) or based on reconstructed signals which are available to the receiver. The exact definition of the combination rule depends on the level of complexity allowed for the module. At the high end, one may compute a possibly complex, optimal predicted frame given all the available information. The enhancement layer residual 116, $r_e(n)$, which is the difference between the input frame 118, $x(n)$, and the predicted frame 104, $\tilde{x}_e(n)$, is then compressed by a compressor 120 to produce the enhancement bits 122.

Referring to FIG. 3 through FIG. 6, a complete scalable predictive coding system for use with this invention is shown. While only three layers are shown, it will be appreciated that additional layers can be added and are contemplated within the scope of the invention. FIG. 3 shows a base layer reconstruction module 124 which receives the quantized base layer residual 112, $\hat{r}_b(n)$, and adds it to the base predicted frame 126, $\tilde{x}_b(n)$, to produce the base layer reconstruction frame 106, $\hat{x}_b(n)$. A delay 128 produces a delayed base reconstructed frame 130, $\hat{x}_b(n-1)$, which is input to the base predictor 132 which computes the base

respectively), and the layer's predicted frame ($\tilde{x}_b(n)$, (EL1) $\tilde{x}_e(n)$, and (EL2) $\tilde{x}_e(n)$, for the base, first enhancement and second enhancement layers, respectively).

The corresponding three layer scalable predictive decoder is shown in FIG. 6. Each layer's inverse compressor/quantizer receives as input the layer's bit stream from which it reproduces the layer's quantized residual. It also extracts the layer's compression parameters for use by a higher layer reconstruction module. The rest of the diagram is identical to the encoder of FIG. 2 and similarly produces the reconstructed frame at each layer.

It will be appreciated that the invention is *generally* applicable to predictive coding and, in particular, may be applied to known *vector quantizer*–based compression techniques, and known *transform*-based techniques. Further, it is applicable to compression of *speech*, *audio*, and *video* signals. A combining rule employing optimal estimation for scalable compression is described next as an implementation example of the invention.

In typical predictive coding, a number of signal representation coefficients (e.g., vectors of transform coefficients, line spectral frequencies, or vectors of raw signal samples) are extracted per frame and quantized independently. A specific low complexity implementation of the invention consists of optimally combining the information available for predicting the coefficient at an enhancement-layer. The reconstructed coefficient at the base-layer, $\hat{x}_b(n)$, and the quantization interval (or partition region in the case of vector quantization) of the corresponding reconstructed residual $\hat{r}_b(n)$, determine an interval/cell $I(n)$ within which the original coefficient $x(n)$ must lie. From the corresponding reconstructed coefficient at the previous enhancement-layer frame, $\hat{x}_e(n-1)$, and a statistical model on time evolution of the coefficients, one may construct a probability density function for $x(n)$ conditional on $\hat{x}_e(n-1)$, denoted by $p[x(n)|\hat{x}_e(n-1)]$. The optimal estimate of $x(n)$ is obtained by expectation:

$$\tilde{x}_e(n) = \frac{\int_{I(n)} xp[x(n)|\hat{x}_e(n-1)]dx}{\int_{I(n)} p[x(n)|\hat{x}_e(n-1)]dx}.$$

This predictor incorporates the information provided by the base-layer (interval within which $x(n)$ lies), and by the enhancement-layer (probability distribution of

hence $r_b(n) \in [a,b]$. Thus the information the base layer provides on *x(n)* is captured in the statement:

$$x(n) \in [\tilde{x}_b(n)+a, \tilde{x}_b(n)+b].$$

At the enhancement layer, the prediction exploits the information available from both layers. The optimal predictor is given therefore by the expectation:

$$\tilde{x}_e(n) = E\{x(n)|\hat{x}_e(n-1), x(n) \in [\tilde{x}_b(n)+a, \tilde{x}_b(n)+b]\},$$

which is conveniently rewritten as

$$\tilde{x}_e(n) = \bar{x}_e(n-1) + E\{z(n)|z(n) \in I_z(n)\}$$

where

$$\bar{x}_e(n-1) = MC[\hat{x}_e(n-1)]$$

and the expectation interval is

$$I_z(n) = [\tilde{x}_b(n)+a-\bar{x}_e(n-1), \tilde{x}_b(n)+b-\bar{x}_e(n-1)].$$

This prediction is directly implemented using the model for *p(z)* given above:

$$\tilde{x}_e(n) = \bar{x}_e(n-1) + \frac{\int_{I_z(n)} zp(z)dz}{\int_{I_z(n)} p(z)dz}.$$

The integral may be analytically evaluated and its closed form solution given explicitly in terms of the integral limits and the parameters α, β, is normally used for simple implementation.

This embodiment of the invention is of low complexity, uses standard video compression for its base layer, and provides substantial performance gains which build up and increase with the number of layers implemented. Its absence in all leading standards in spite of its gains and low complexity strongly suggests that the invention is not obvious to the leading researchers and developers in the field of video compression.

The scalable predictive coding method of the invention, although illustrated herein on a two or three-layer scalable system, is repeatedly applicable to further layers of enhancement in a straightforward manner. For example, at layer *k* we combine signal information from the current reconstructed frame at layer *k-1*, and from the previous reconstruction frame at layer *k*. A higher complexity version allows for the combining rule to take into account data from all lower layers. In the special implementation described, information from all lower layers contributes to

CLAIMS

What is claimed is:

1. A method for predicting the current frame of data in a digital coding system wherein a signal is segmented into frames of data that are sequentially encoded, said system including a base layer and an enhancement layer, said base layer including a base encoder and a base decoder, said enhancement layer including an enhancement encoder and an enhancement decoder, said base decoder producing a reconstructed signal, said enhancement decoder producing an enhanced reconstructed signal, said method comprising the steps of:

predicting the current frame of data at the enhancement-layer by processing and combining the reconstructed data representing the current base layer frame and the reconstructed data representing the previous enhancement layer frame.

2. A method for scalable predictive coding of a signal, comprising the steps of:

(a) encoding data representing said signal with a base layer predictive coding system that provides a first prediction of said signal and information indicative of a decoded base layer approximation to said signal;

(b) encoding data representing said signal by a first enhancement layer which performs predictive coding with a second prediction of said signal derived from a combination of information from the base layer and information indicative of the past decoded signal approximation generated in said first enhancement layer.

3. A method as recited in claim 2, wherein the step of encoding said signal data with said enhancement layer comprises the steps of providing to said first enhancement layer compression parameters from the base layer to aid in the computation of said second prediction.

4. A method as recited in claim 2, wherein the step of encoding said signal data with said first enhancement layer comprises the steps of providing to said first enhancement layer time evolution statistics derived ither by off-line computation or by computations using quantized parameters of said signal.

9. An apparatus as recited in claim 8, wherein said means for encoding said signal data with said enhancement layer comprises means for providing to said first enhancement layer compression parameters from the base layer to aid in the computation of said second prediction.

10. An apparatus as recited in claim 8, wherein said means for encoding said signal data with said first enhancement layer comprises means for providing to said first enhancement layer time evolution statistics derived either by off-line computation or by computations using quantized parameters of said signal.

11. An apparatus as recited in claim 8, further comprising a second enhancement layer, wherein said second enhancement layer performs predictive coding with a third prediction of said signal derived from a combination of information from said first enhancement layer and information indicative of the past decoded signal approximation generated in said second enhancement layer.

12. An apparatus as recited in claim 8, wherein said second prediction at predetermined intervals is derived exclusively from information from the base layer and at all other times is derived by combining information from the base layer and information indicative of the past decoded signal approximation generated in said first enhancement layer.

13. A scalable predictive coding system for compressing a signal, comprising at least one enhancement layer and at least one lower layer, wherein prediction in an enhancement layer combines information from a lower layer with information from the enhancement layer.

14. A scalable predictive coding method for compressing a signal in system comprising at least one enhancement layer and at least one lower layer, the method comprising the steps of performing prediction in an enhancement layer by combining information from a lower layer with information from the enhancement layer.

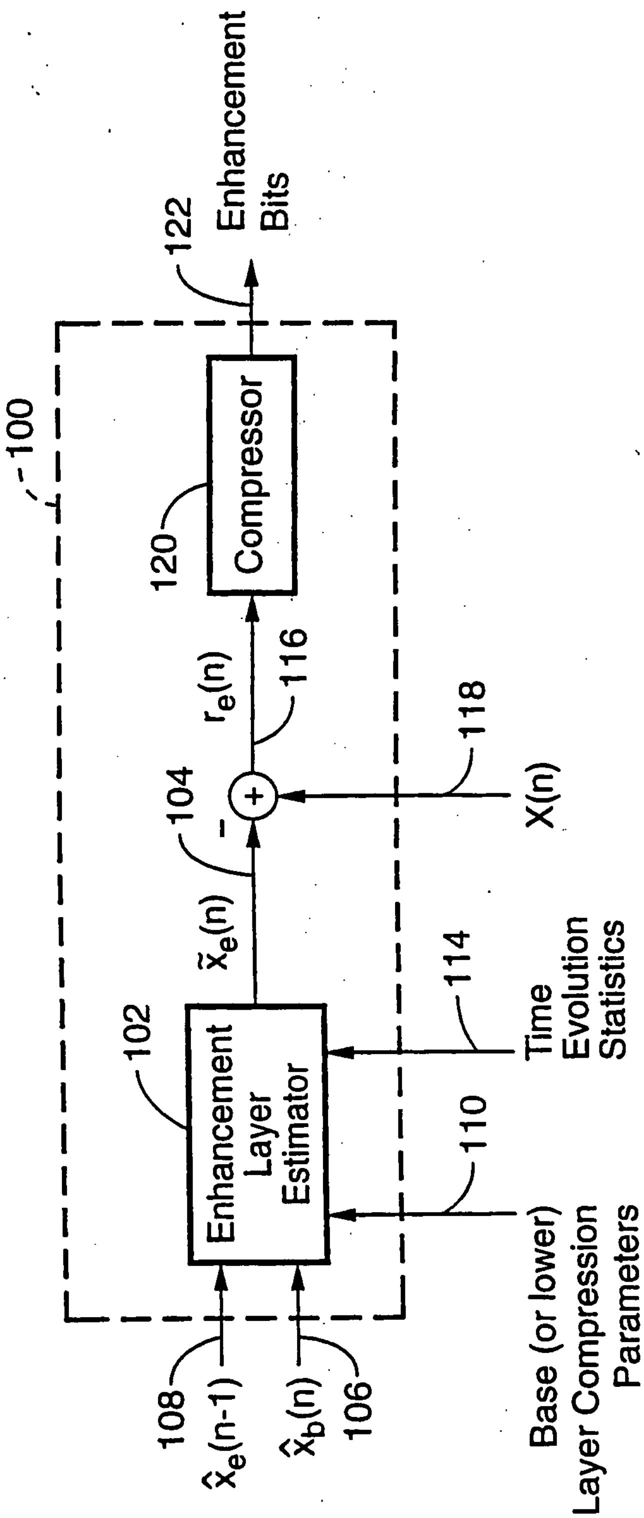

100
102
Enhancement Layer Estimator
120
Compressor
122
Enhancement Bits
104
+
-
116
$r_e(n)$
$\tilde{x}_e(n)$
118
X(n)
114
Time Evolution Statistics
110
Base (or lower) Layer Compression Parameters
108
$\hat{x}_e(n\text{-}1)$
$\hat{x}_b(n)$
106

FIG. – 2

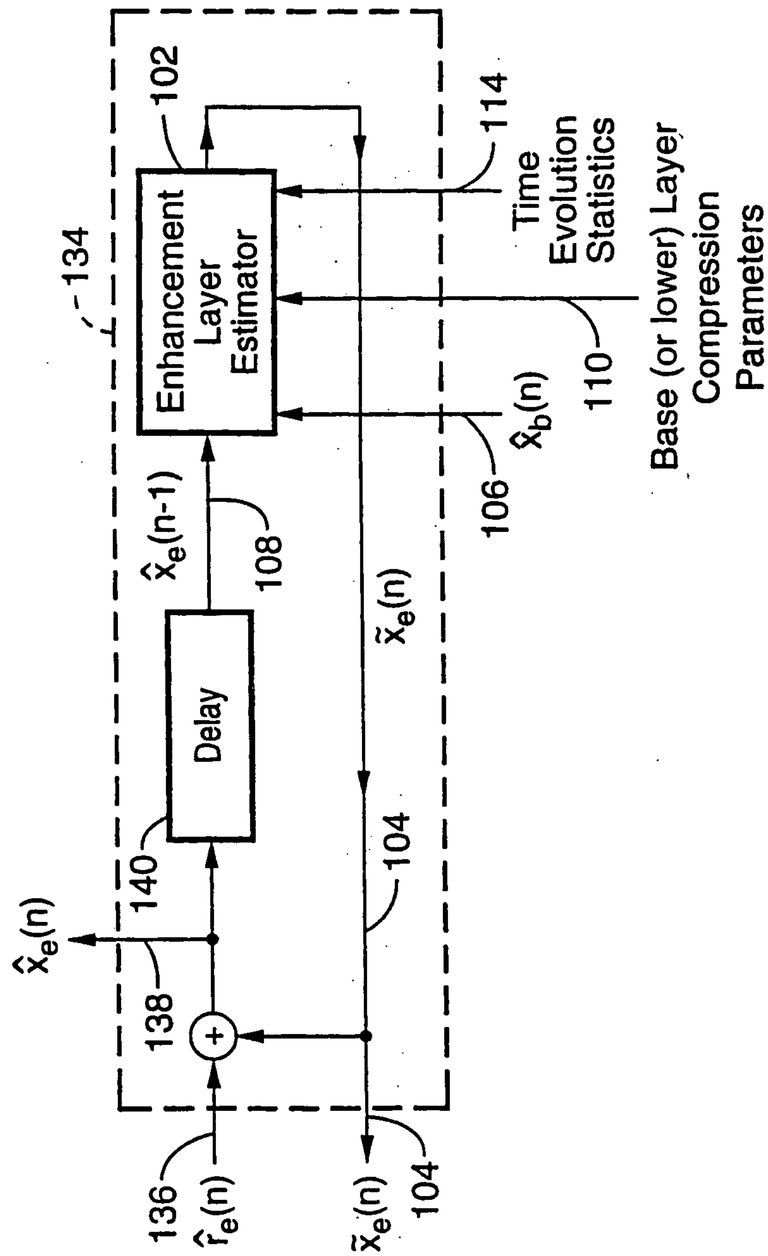

Enhancement Layer Estimator
Delay
102
114
Time Evolution Statistics
134
Base (or lower) Layer Compression Parameters
110
$\hat{x}_b(n)$
106
$\hat{x}_e(n-1)$
108
$\tilde{x}_e(n)$
104
140
$\hat{x}_e(n)$
138
136
$\hat{r}_e(n)$
$\tilde{x}_e(n)$
104

FIG. – 4

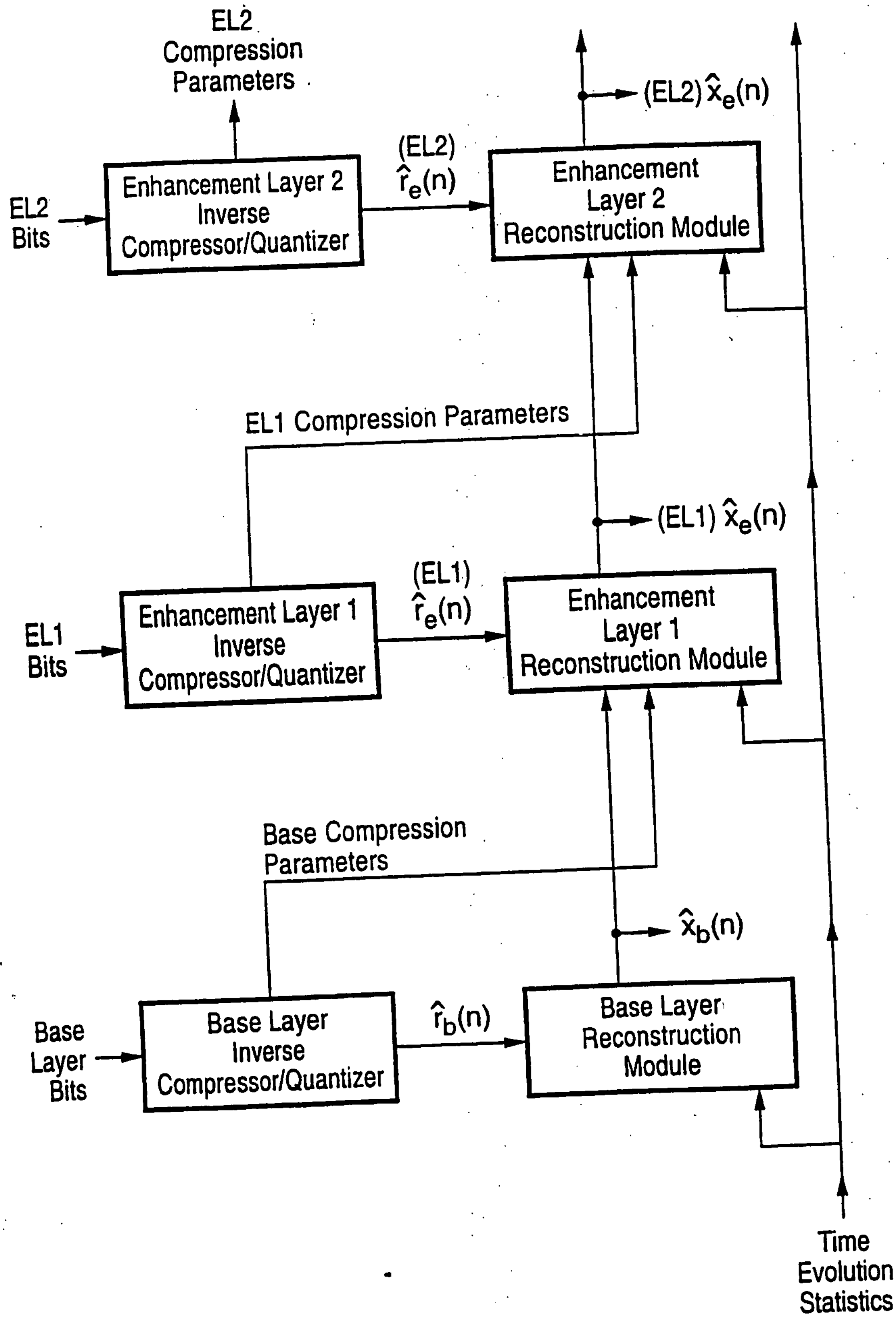

EL2 Compression Parameters
EL2 Bits
Enhancement Layer 2 Inverse Compressor/Quantizer
(EL2) $\hat{r}_e(n)$
Enhancement Layer 2 Reconstruction Module
(EL2) $\hat{x}_e(n)$
EL1 Compression Parameters
(EL1) $\hat{x}_e(n)$
EL1 Bits
Enhancement Layer 1 Inverse Compressor/Quantizer
(EL1) $\hat{r}_e(n)$
Enhancement Layer 1 Reconstruction Module
Base Compression Parameters
$\hat{x}_b(n)$
Base Layer Bits
Base Layer Inverse Compressor/Quantizer
$\hat{r}_b(n)$
Base Layer Reconstruction Module
Time Evolution Statistics

FIG. – 6

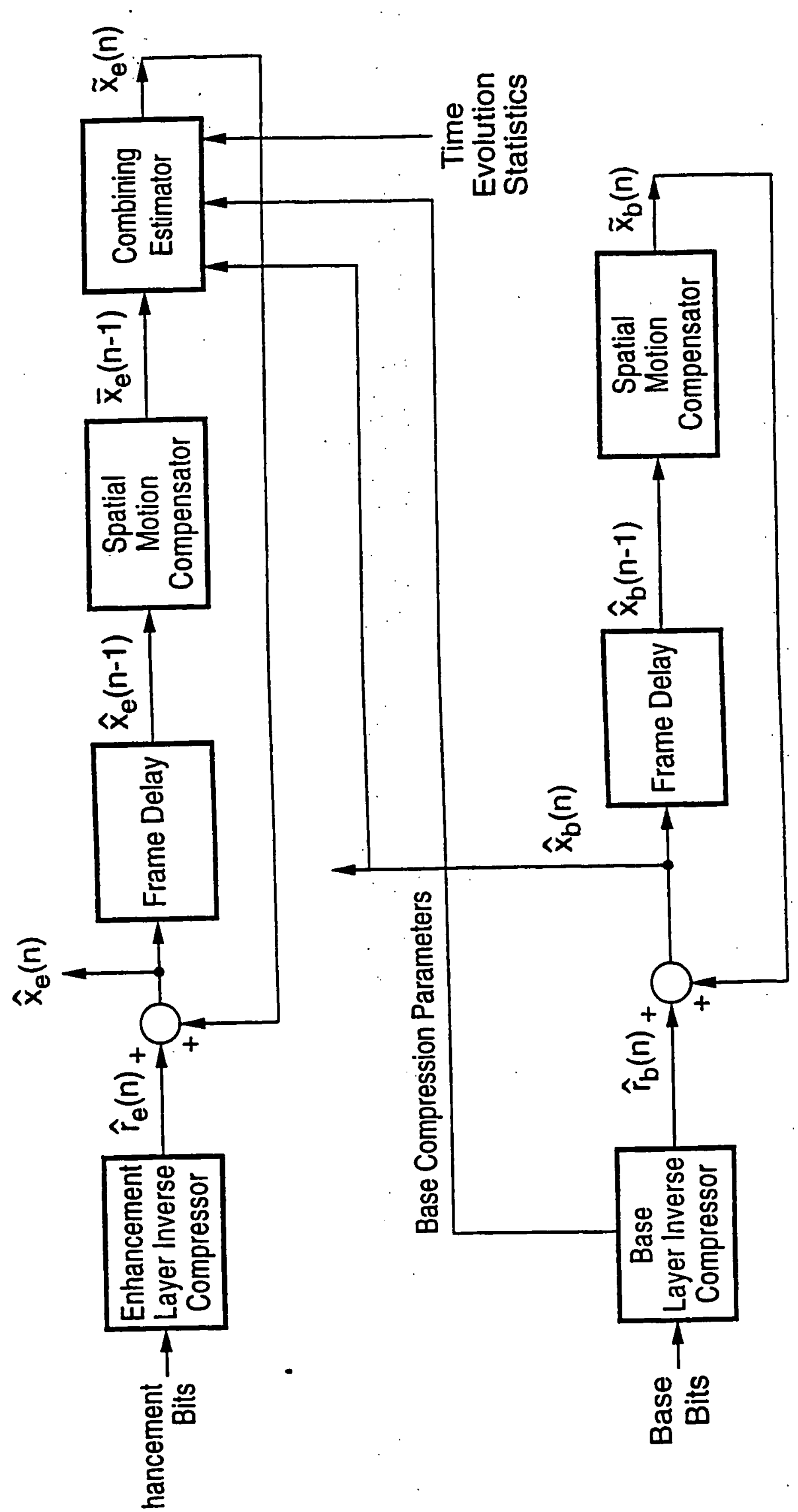

hancement Bits
Enhancement Layer Inverse Compressor
$\hat{r}_e(n)$
+
+
$\hat{x}_e(n)$
Frame Delay
$\hat{x}_e(n-1)$
Spatial Motion Compensator
$\bar{x}_e(n-1)$
Combining Estimator
$\tilde{x}_e(n)$
Time Evolution Statistics
Base Compression Parameters
$\hat{x}_b(n)$
Base Bits
Base Layer Inverse Compressor
$\hat{r}_b(n)$
+
+
Frame Delay
$\hat{x}_b(n-1)$
Spatial Motion Compensator
$\tilde{x}_b(n)$

FIG. – 8

INTERNATIONAL SEARCH REPORT

Intern. ıal Application No
PCT/US 98/26984

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04N H03M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 634 871 A (AT & T CORP) 18 January 1995 see column 20, line 10 - line 57; figure 20 ---	1,2,7,8, 13,14
X	EP 0 771 119 A (TOKYO SHIBAURA ELECTRIC CO) 2 May 1997 see page 6, line 53 - page 7, line 42; figure 1 ---	13,14
X	EP 0 644 695 A (AT & T CORP ;BELL COMMUNICATIONS RES (US)) 22 March 1995 see column 13, line 26 - column 15, line 40; figure 12 ---	13,14
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 April 1999	21/04/1999

Name and mailing address of the ISA	Authorized officer

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter   onal Application No
PCT/US 98/26984

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0634871 A	18-01-1995	CA 2126467 A	14-01-1995
		JP 7170514 A	04-07-1995
		US 5742343 A	21-04-1998
EP 0771119 A	02-05-1997	JP 9182084 A	11-07-1997
		US 5818531 A	06-10-1998
EP 0644695 A	22-03-1995	CA 2127151 A	22-03-1995
		JP 7162870 A	23-06-1995

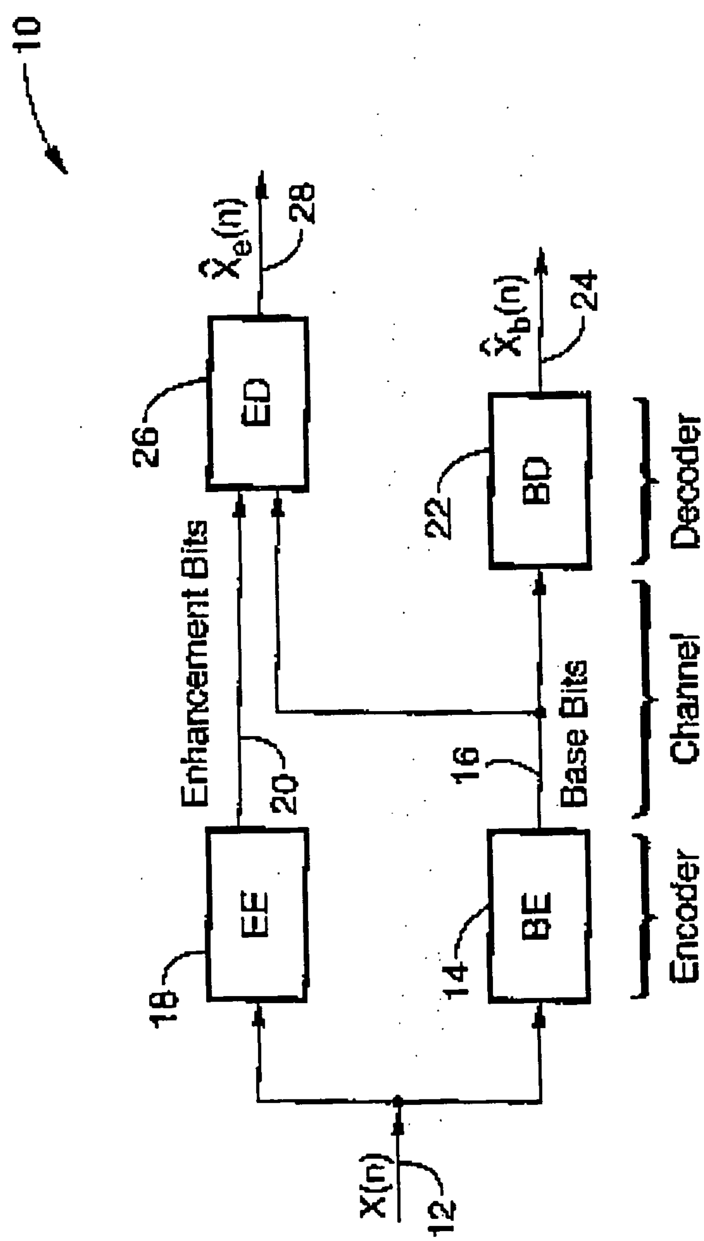


FIG. – 1
(PRIOR ART)

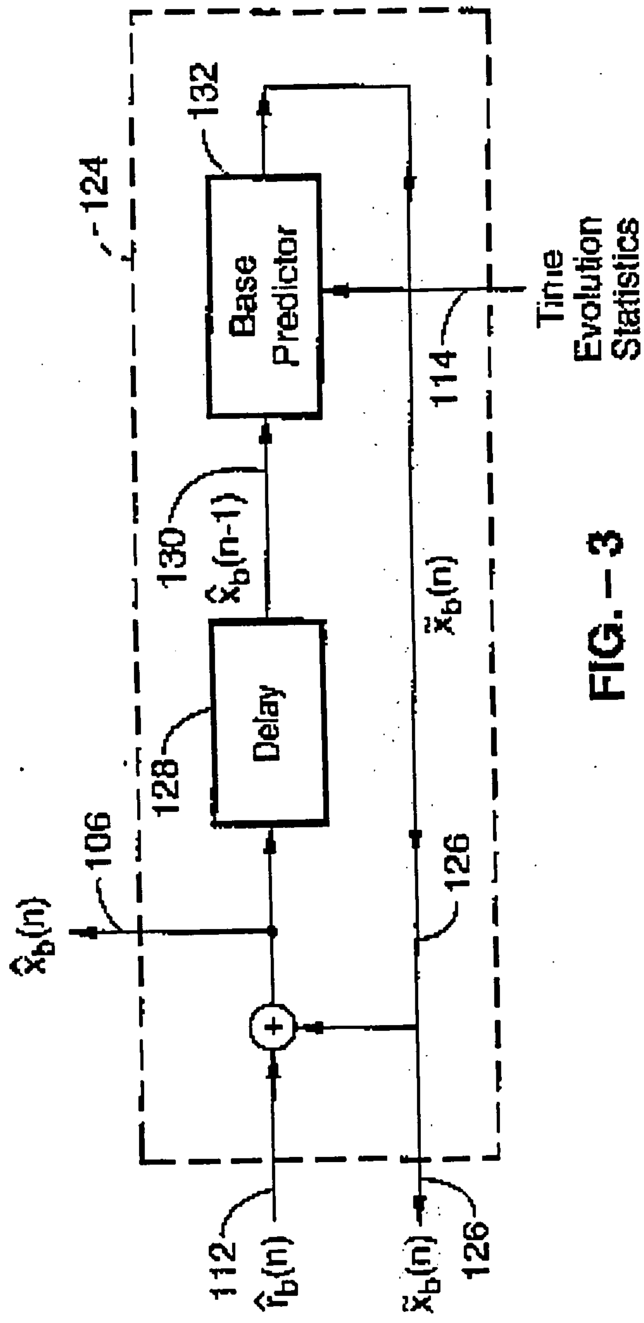

124
Base Predictor
132
Delay
128
130
106
126
112
114
Time Evolution Statistics
$\hat{x}_b(n)$
$\hat{x}_b(n-1)$
$\tilde{x}_b(n)$
$\hat{r}_b(n)$
+

FIG. — 3

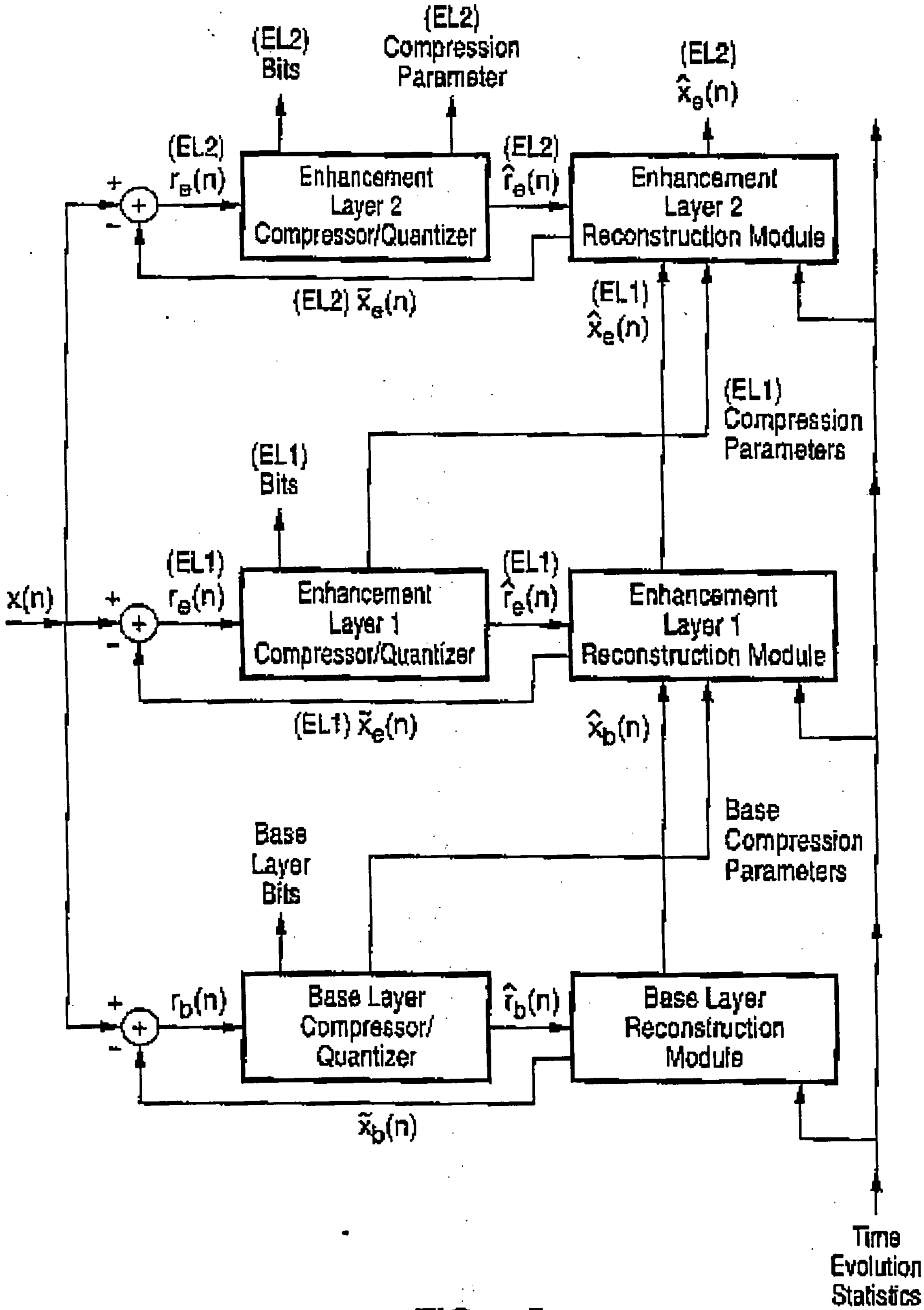

(EL2) Bits
(EL2) Compression Parameter
(EL2) $\hat{x}_e(n)$
(EL2) $r_e(n)$
Enhancement Layer 2 Compressor/Quantizer
(EL2) $\hat{r}_e(n)$
Enhancement Layer 2 Reconstruction Module
(EL2) $\tilde{x}_e(n)$
(EL1) $\hat{x}_e(n)$
(EL1) Compression Parameters
(EL1) Bits
x(n)
(EL1) $r_e(n)$
Enhancement Layer 1 Compressor/Quantizer
(EL1) $\hat{r}_e(n)$
Enhancement Layer 1 Reconstruction Module
(EL1) $\tilde{x}_e(n)$
$\hat{x}_b(n)$
Base Compression Parameters
Base Layer Bits
$r_b(n)$
Base Layer Compressor/ Quantizer
$\hat{r}_b(n)$
Base Layer Reconstruction Module
$\tilde{x}_b(n)$
Time Evolution Statistics

FIG. – 5

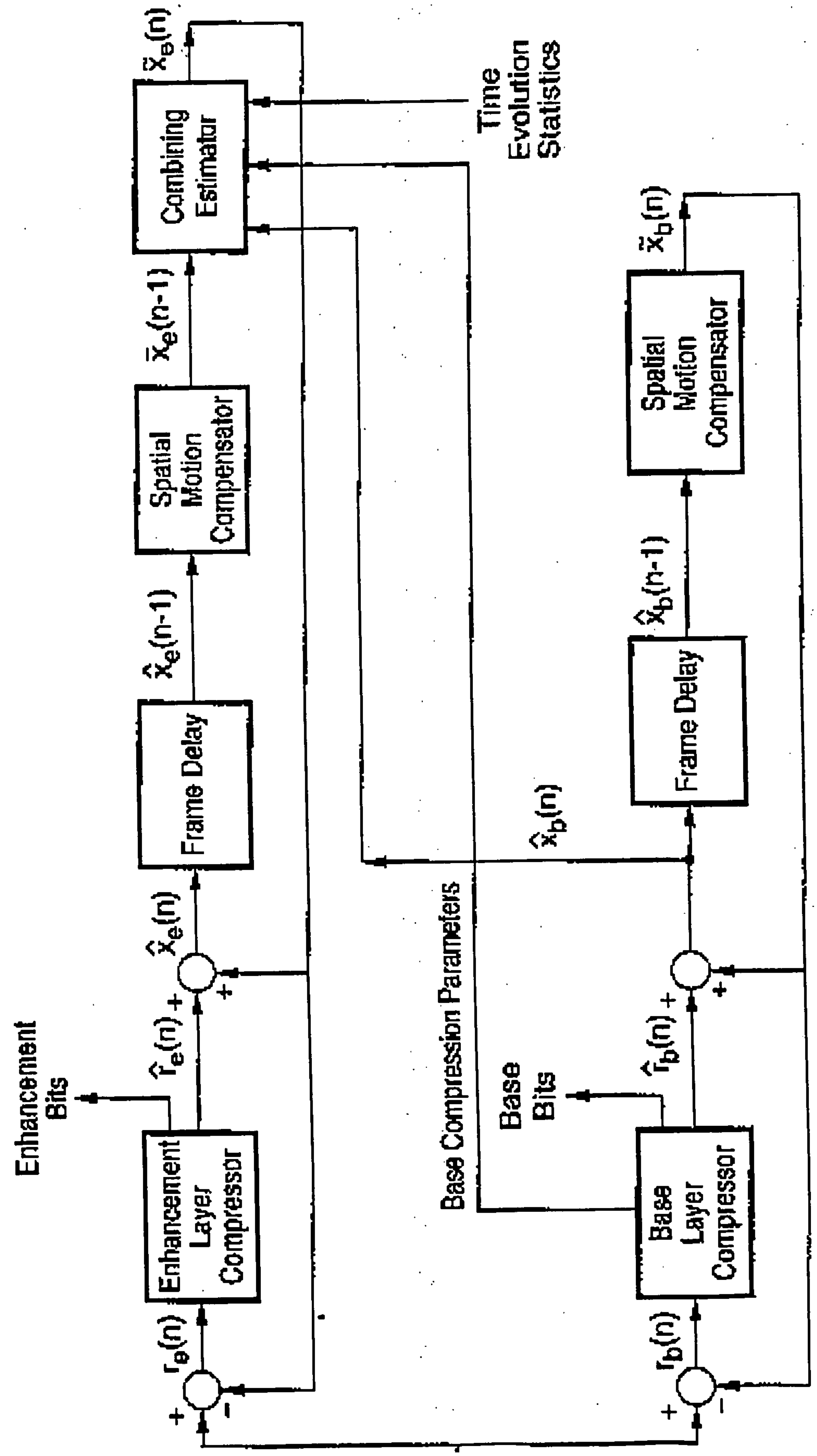

Enhancement Layer Compressor
Enhancement Bits
Frame Delay
Spatial Motion Compensator
Combining Estimator
Time Evolution Statistics
Base Compression Parameters
Base Layer Compressor
Base Bits
Frame Delay
Spatial Motion Compensator

FIG. - 7

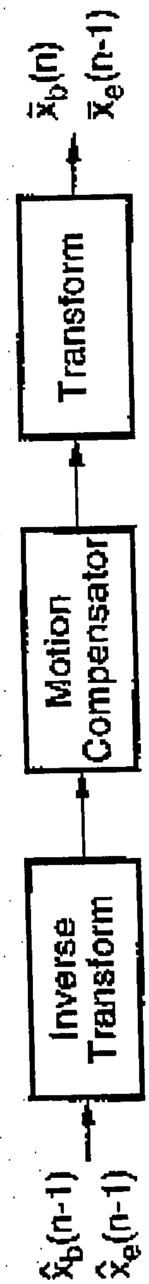

$\hat{x}_b(n-1)$
$\hat{x}_e(n-1)$
Inverse Transform
Motion Compensator
Transform
$\bar{x}_b(n)$
$\bar{x}_e(n-1)$

FIG. — 9